# A Perspective Projection Camera Model
# for Zoom Lenses

R. G. Willson  and  S. A. Shafer

Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania  15213
U.S.A.

## Abstract

To effectively use automated zoom lenses for machine vision we need camera models that are valid over continuous ranges of lens settings. While camera calibration has been the subject of much research in machine vision and photogrammetry, for the most part the resulting models and calibration techniques have been for cameras with fixed parameter lenses where the lens' imaging process is static. For cameras with automated lenses the image formation process is a dynamic function of the lens control parameters. The complex nature of the relationships between the control parameters and the imaging process plus the need to calibrate them over a continuum of lens settings makes both the modeling and the calibration of cameras with automated zoom lenses fundamentally more difficult than that of cameras with fixed parameter lenses.

In this paper we illustrate some of the problems involved with the modeling and calibration of cameras with variable parameter lenses. We then show how an iterative, empirical approach to modeling and calibration can produce a dynamic camera model of perspective projection that holds calibration across a continuous range of zoom.

## 1  Introduction

Camera systems with automated zoom lenses are inherently more useful than those with fixed parameter lenses. With a variable parameter lens a camera can adapt to changes or differences in the scenes being imaged, focus attention on specific objects in view that differ in size and location, or even measure properties of the scene by noting how the image changes as the lens' parameters are varied. But, to effectively use

zoom lenses for machine vision we need camera models that are valid over continuous ranges of lens settings.

## 1.1   Fixed vs Variable Parameter Lenses

In modeling and calibrating automated zoom lenses our end objective is to capture the net relationship between the lens control parameters and some aspect of the image formation process. Conceptually this relationship can be subdivided into two parts, as illustrated in Fig. 1. The first part, $R_1$, is the relationship between the image formation process and hardware configuration of the lens. The hardware configuration is specified by the composition, dimensions and positions of the optical components of the lens. The second part, $R_2$, is between the hardware configuration of the lens and lens' control parameters (if any). In fixed parameter camera systems the lens' hardware configuration is static and we need to consider only $R_1$ in the modeling and calibration of the lens. In variable parameter camera systems the lens' hardware configuration is dynamic and we must consider both $R_1$ and $R_2$.

## $R_1$ - Hardware Configuration and the Image Formation Process

For real lenses the low level optics relating the lens' hardware configuration to the actual image formation process is generally too complex to be expressed in closed-form equations, even for a simple fixed parameter camera lens. Lens designers deal with this by resorting to simulations of the image formation process using ray tracing [1]. In ray tracing the paths of individual light rays are traced as they refract at each optical surface in the lens. With enough rays the designer can characterize the lens' image formation process sufficiently to evaluate the lens' design. While the equations used in ray tracing are explicitly related to the hardware configuration of the lens, they cannot be used to build parameterized models of the imaging properties that we are interested in.

In machine vision we are interested in higher level aggregate properties of the image formation process. These range from simple image properties such as magnification and focussed distance, to more complex image properties such as perspective projection and amount of image defocus. In order to have computationally efficient closed-form equations for these properties the models must necessarily be based on simplifications or abstractions of the actual image formation process. The two most common abstract models are the pinhole camera model and the thin-lens camera model, used respectively to explain perspective projection and image defocus.

## $R_2$ - Control Parameters and the Hardware Configuration

In lenses the relationship between the lens' control parameters and the actual hardware configuration of the lens is essentially an arbitrary design choice made by the manufacturer. Typically the relationship is hidden from the user. Worse still, the mechanical nature of this relationship introduces several difficult modeling and calibration problems. To illustrate these problems we can look at the position of the image of a laser (initially autocollimated at one specific lens setting to determine the optical axis) as the focus and zoom of a precision automated zoom lens are varied. In Fig. 2 we can see significant hysteresis in the position of the laser's image as the focus control is varied from 1.0 m to $\infty$ and back to 1.0 m. Figure 3 shows both hysteresis and a sharp discontinuity in the laser's position as the zoom control is
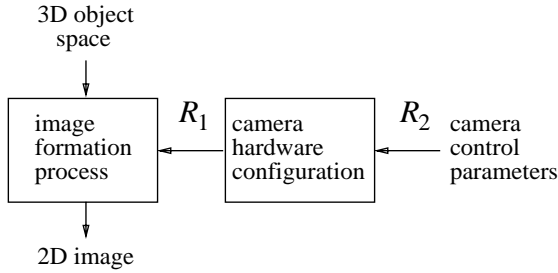
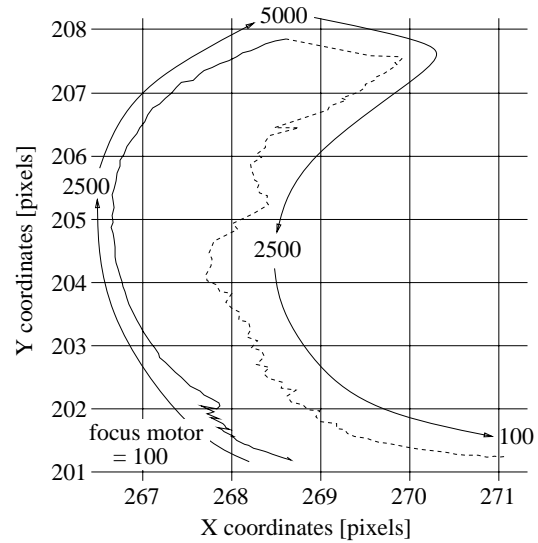Figure 1: Lens control parameters and the image formation process



Figure 2: Hysteresis in position of laser's image during focusing

varied from 130 mm to 10 mm. In both these examples the automation for the lens is provided by highly repeatable digital microstepping motors (see Section 2), thus the error is due primarily to the internal mechanical and optical properties of the lens. Notwithstanding mechanical hysteresis and discontinuities, with precise automation and control the hardware configuration of the lens can be made to be very repeatable and thus calibratable. Figure 4 shows the repeatability of the position of the laser's image as the focus is varied twice from 1.0 m to $\infty$.

## 1.2 Modeling Variable Parameter Lenses

For fixed parameter lenses the image formation process is static and thus the terms in our camera models are constants. In variable parameter lenses the image formation process is a dynamic function of the lens control parameters, and thus the terms of our camera models must also be variable. The question is, How do the terms vary with the control parameters? This is a difficult question to answer for two reasons. First, the two traditional models of the image formation process, the pinhole camera and the thin-lens, are idealized high level abstractions of the real image formation process and the connection between the lens' physical configuration and the model terms is not direct. Second, as we've seen the relationship between the lens' physical configuration and the control parameters is complex and typically unknown. The answer to the question then is that we have no good theoretical basis for the relationships between the terms of our camera models and the lens control parameters. Every model term is potentially a function of every lens control parameter [5]. The actual relationships must be determined empirically.

The most direct method of discovering the relationships between the model terms and the lens control parameters would be to step the lens through the full range of its control parameters while performing a full calibration of the camera model at each
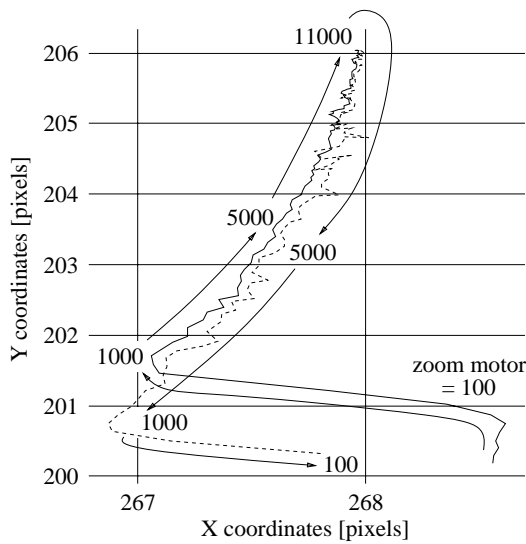
Figure 3: Discontinuity and hysteresis in position of laser's image during zooming
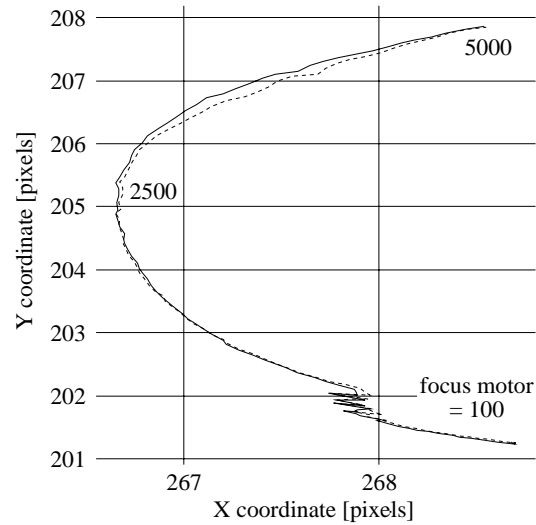
Figure 4: Repeatability of position of laser's image over focus

step[1]. The drawback with this method is the amount of computational effort required to determine all of the model terms at every lens setting.

While every model term is potentially a function of every lens control parameter, in reality the dependencies between the model terms and the control parameters range from strong, to weak, to none at all. By taking advantage of the fact that some model terms remain relatively constant over ranges of the lens control parameters we can greatly reduce the effort required to determine the variation in the remaining model terms.

### 1.3   Calibrating Variable Parameter Lenses

Unlike the calibration of fixed parameter lenses, the calibration of variable parameter lenses requires that measurements be made over ranges of hardware configurations for the lens. This raises several challenges. First, the dimensionality of the data is the same as the number of control parameters that are to be concurrently modeled. Ten measurements across the range of each of the focus, zoom and aperture controls gives us 1000 hardware configurations to calibrate for, compared to just one for a fixed parameter lens system.

A second challenge is the potential difficulty in taking measurements across the wide range of imaging situations that can occur over the range of some control parameters. As an example, consider the measurement of features on a calibration target as the zoom is varied. As the lens is zoomed in (i.e. the focal length is increased) the number of feature points in the camera's field of view may decrease below the number necessary to perform an accurate calibration. Conversely, as the lens is zoomed out the features on the target may become too small and/or crowded to be accurately measured. In the end, several targets with different scales may be required to cover

---

[1]This was the approach used by Wiley [4] on a manually adjusted lens.

the full range of zoom. Taking measurements over wide ranges of focus and defocus can also be problematic.

When collecting the calibration data the sampling interval(s) depend on the rate of variation of the model terms with each control parameter. An initial sparse sampling along with full camera calibration can be used to identify the rate and degree of variation of all of the model terms with respect to a given control parameter. Slowly varying or non varying terms can then be modeled and calibrated using the sparse calibration data. Where terms vary rapidly, denser calibration data is taken. Using this approach the computation required to model and calibrate the more rapidly varying terms is reduced by already having models (and thus values) for the more slowly varying terms.

## 2  A Perspective Projection Camera Model for Zoom Lenses

In this section we use the above techniques to develop a perspective projection camera model that "holds calibration" across a continuous range of the zoom control parameter. By "holds calibration" we mean that the average magnitude of the error in the image plane between the calibrated camera model and calibration data remains essentially constant across the desired range of zoom.

The camera system for this work consists of a Fujinon A13$\times$10BRM-8 zoom lens mounted on a Photometrics Star 1 digital camera. Automation for the lens is provided by digital microstepping motors which are connected to the lens body by backlash-free pushrod and pulley assemblies [5]. The microstepping motors provide repeatable drift free positioning of the lens hardware, even across powerdowns. The lens has 5100 steps of resolution for focus, 11100 steps for zoom, and 2700 steps for aperture.

The calibration target is a white plane containing 3.2 mm diameter black dots on regular 12.7 mm grid. The target is mounted parallel to the camera's sensor plane on a linear positioner who's axis is parallel to the camera's optical axis.

The calibration data that we use to develop our camera model was taken over 31 lens settings, from zoom motor ($m_z$) positions 1800 to 2100 in 10 unit steps. At each lens setting images of the calibration target were taken at ranges of 2800 mm, 2550 mm and 2300 mm between the target plane and the camera's sensor plane. Each image contained between 272 and 460 control points, depending on the zoom setting and on the range to the target.

### 2.1  The Static Camera Model

To model the image formation process for any given lens setting we use the 11 term pinhole camera model described in Tsai [3]. As illustrated in Fig. 5, the origin of the camera-centered coordinate system $(x_c, y_c, z_c)$ coincides with the front nodal point of the camera and the $z_c$ axis coincides with the camera's optical axis. The image plane is assumed to be parallel to the $(x_c, y_c)$ plane at a distance $f$ from the origin, where $f$ is the effective focal length of the camera.

The relationship of between the position of a point $P$ in world coordinates, $(x_w, y_w, z_w)$, and the point's image in the camera's frame buffer, $(X_f, Y_f)$, is defined by a sequence of coordinate transformations. The first is a rigid body transformation from the world coordinate system $(x_w, y_w, z_w)$ to the camera-centered coordinate system $(x_c, y_c, z_c)$.
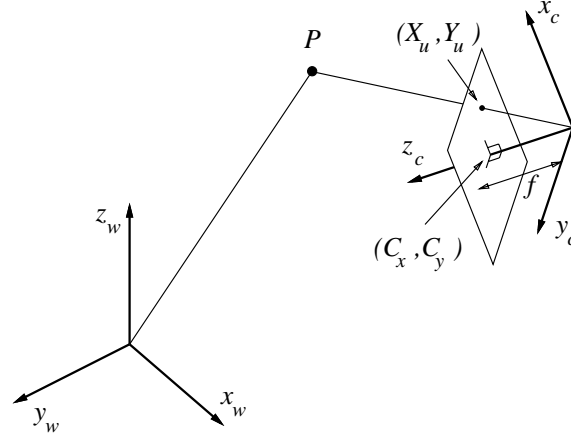
Figure 5: Camera model geometry

This is expressed as

$$
\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \quad \text{where} \quad R = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix} \tag{1}
$$

is the $3 \times 3$ rotation matrix describing the orientation of the camera in the world coordinate system.

The second transformation is a perspective projection (using an ideal pinhole camera model) of the point in camera coordinates to the position of its image in undistorted sensor plane coordinates, $(X_u, Y_u)$. This is described by

$$
X_u = f\frac{x_c}{z_c} \quad \text{and} \quad Y_u = f\frac{y_c}{z_c} \tag{2}
$$

where $f$ is the effective focal length of the pinhole camera.

The third transformation is from the undistorted (ideal) position of the point's image in the sensor plane to the true position of the point's image, $(X_d, Y_d)$, that results from geometric lens distortion. This is described by

$$
X_u = X_d(1 + \kappa_1 \rho^2), \quad Y_u = Y_d(1 + \kappa_1 \rho^2) \quad \text{and} \quad \rho = \sqrt{X_d^2 + Y_d^2}
$$

where $\kappa_1$ is the coefficient of radial lens distortion. While a more complex model describing both radial and tangential geometric lens distortion could have been used, the accuracy provided by this model is sufficient to demonstrate the development of our zoom lens model.

The final transformation in the static camera model is between the true position of the point's image on the sensor plane and its coordinates in the camera's frame buffer, $(X_f, Y_f)$. This is described by

$$
X_f = d_x^{-1} X_d s_x + C_x \quad \text{and} \quad Y_f = d_y^{-1} Y_d + C_y
$$

where $C_x$ and $C_y$ are the coordinates (in pixels) of the intersection of the $z_c$ axis and the camera's sensor plane, $d_x$ and $d_y$ are the effective center to center distances between

the camera's sensor elements in the $x_c$ and $y_c$ directions, and $s_x$ is a scaling factor to compensate for any uncertainty in the ratio between the number of sensor elements on the CCD and the number of pixels in the camera in the $x$ direction.

## 2.2 Iterative Development of the Dynamic Camera Model

We start the development of the dynamic camera model by first fully calibrating the static camera model at one zoom position ($m_z = 1800$). The full calibration is done using Tsai's algorithm [3] to generate an initial set of values for the model's 11 terms followed by a general non-linear optimization to further refine them.

From the static camera model we can see that the primary effect of the lens' zoom control will be on the focal length term, $f$. Starting from the base set of constants obtained at $m_z = 1800$, we estimate $f$ from the calibration data for $m_z = 1810\ldots2100$. Figure 6 shows $f$ versus $m_z$. If we look at the average magnitude of the image plane error between the model and the calibration data (top curve in Fig. 7) we see that it climbs by an order of magnitude across the chosen range of zoom motor settings. Clearly not all of the dependencies between the model terms and $m_z$ have been captured by allowing just the $f$ term of the model to vary.

A second order effect of changing the zoom is a shifting of the camera's field of view due to changes in the optical alignment of the lens components [6]. In the camera model this shifting of the field of view can be accommodated by letting $C_x$ and $C_y$ vary with $m_z$. Starting again from the base set of constants at $m_z = 1800$, this time we estimate $f$, $C_x$ and $C_y$ from the calibration data for $m_z = 1810\ldots2100$. Figures 8 and 9 show $C_x$ and $C_y$ versus $m_z$. Looking again at the mean image plane error (middle curve in Fig. 7) this time we see that it climbs by only a factor of two across the range of zoom motor settings. While this represents a significant improvement over the first attempt, clearly there are more dependencies between the model terms and $m_z$.

Another second order effect of changing zoom is a shifting of the position of the camera's nodal points due to the repositioning of lens components along the optical axis [2]. In the camera model this shifting is equivalent to a change in the position of the origin in the camera's coordinate frame along the camera's $z$ axis, and can be accommodated by allowing the model's $T_z$ term to vary with $m_z$. Starting again from the base set of constants at $m_z = 1800$, this time we estimate $f$, $C_x$, $C_y$ and $T_z$ from the calibration data for $m_z = 1810\ldots2100$. Figure 10 shows $T_z$ versus $m_z$. Looking at the mean image plane error between the model and the calibration data (bottom curve in Fig. 7) we see that this time the error is relatively flat across the range of zoom motor settings.

To get an idea of how well the above camera model is holding calibration we perform a full static camera calibration on the calibration data from each lens setting ($m_z = 1800\ldots2100$). This gives us the minimum mean image plane error possible for this set of calibration data for each given $m_z$. The results indicate that by modeling just four of the 11 camera model terms as functions of the zoom control parameter we can hold the mean image plane error to within 1.5 % of the best error obtainable for the full 11 term static camera calibration across this range of the zoom parameter (if the full static camera calibration error were plotted in Fig. 7 it would completely overlay the bottom curve).
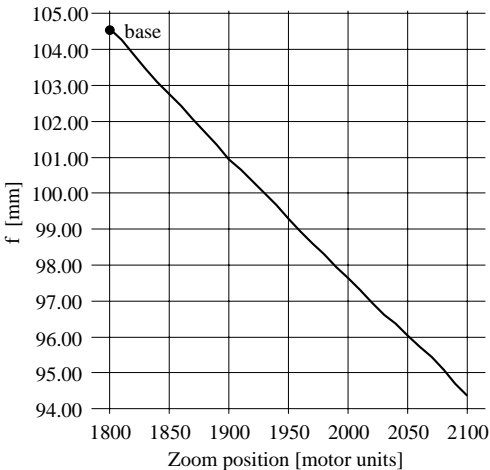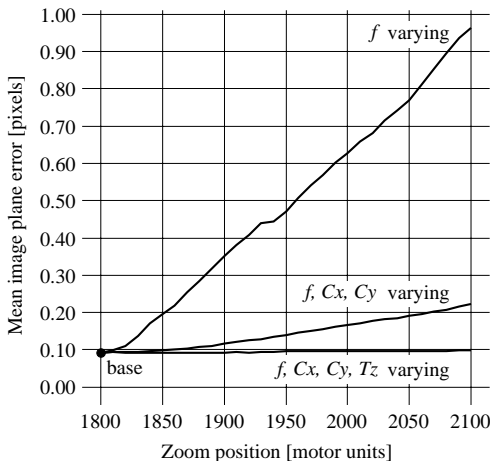
Figure 6: $f$ versus zoom motor

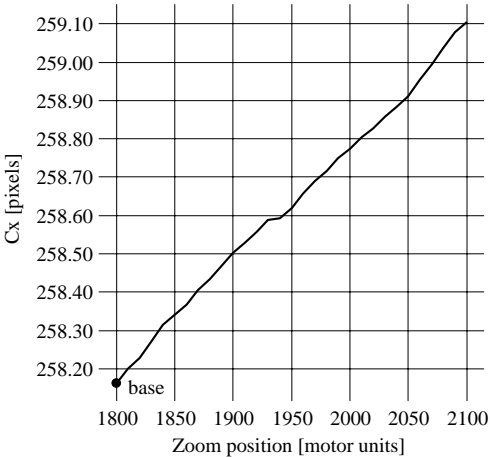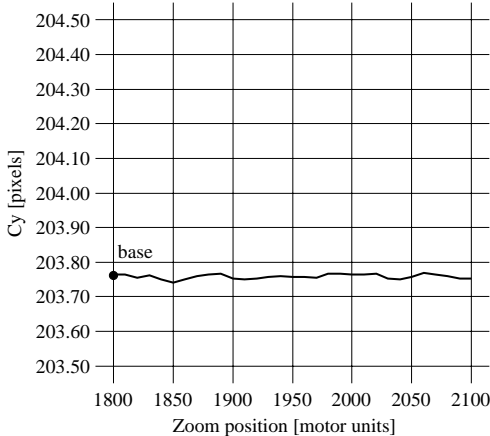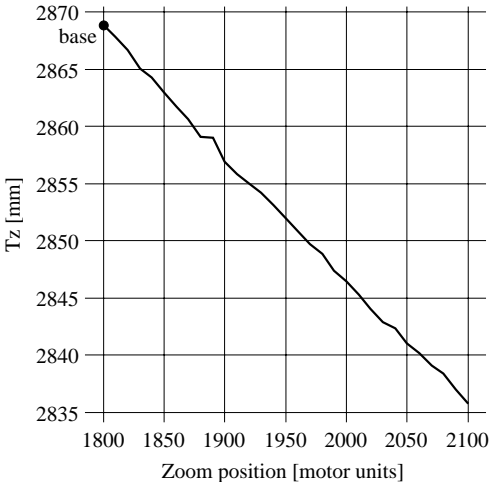Figure 7: Mean image plane error versus zoom motor

Figure 8: $C_x$ versus zoom motor

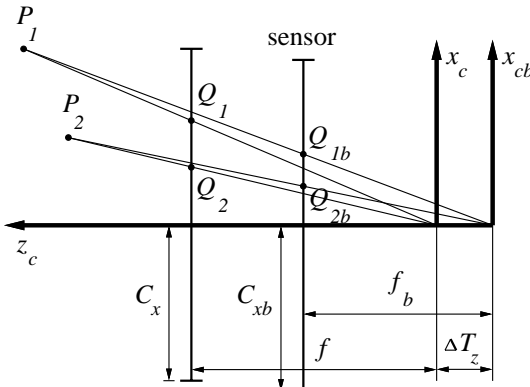Figure 9: $C_y$ versus zoom motor

Figure 10: $T_z$ versus zoom motor

Figure 11: 2D camera geometry with varying $f$, $C_x$, $C_y$ and $T_z$ terms

## 2.3   Efficient Estimation of $f$, $C_x$, $C_y$ and $T_z$ Model Terms

In the previous section we showed that with variations in only four of the 11 model terms our camera model could hold calibration across a continuous range of the zoom motor. Allowing only the $f$, $C_x$, $C_y$ and $T_z$ terms to vary also permits a special reformulation of the equations that makes calibration easier. Figure 11 shows a 2D illustration of the camera's imaging geometry when the $f$, $C_x$, $C_y$ and $T_z$ terms are allowed to vary. At the base lens setting the points $P_1$ and $P_2$ have images $Q_{1b}$ and $Q_{2b}$ on the sensor plane located at a focal length of $f_b$ from the camera's origin. When $m_z$ is changed the effective focal length changes from $f_b$ to $f$, the image center shifts perpendicular to the $z_c$ axis from $C_{xb}$ to $C_x$ and the camera origin shifts along the $z_c$ axis by $\Delta T_z$, causing the images of the two points move to positions $Q_1$ and $Q_2$. Using simple geometry the relationship between $C_x$ and $C_{xb}$ (and similarly between $C_y$ and $C_{yb}$) is

$$C_x = Q_1 - (Q_{1b} - C_{xb}) \frac{|Q_1 - Q_2|}{|Q_{1b} - Q_{2b}|}$$

.

For the new lens setting equations (1) and (2) give us

$$X_{u_i} = f \frac{x_{c_i}}{z_{c_i}} = f \frac{r_1 x_{w_i} + r_2 y_{w_i} + r_3 z_{w_i} + T_x}{r_7 x_{w_i} + r_8 y_{w_i} + r_9 z_{w_i} + T_z} = f \frac{x_{c_i}}{K_{z_i} + T_z}$$

for every point $P_i$, where $K_{z_i} = r_7 x_{w_i} + r_8 y_{w_i} + r_9 z_{w_i}$. Rearranging terms we get

$$f x_{c_i} - T_z X_{u_i} = K_{z_i} X_{u_i}$$

where $x_{c_i}$, $X_{u_i}$ and $K_{z_i}$ can be directly calculated for all points $P_i$ using the calibration data from the new lens setting plus the values of $C_x$ and $C_y$ obtained above. To find $f$ and $T_z$ for the new lens setting we simply solve the over determined set of linear equations

$$
\begin{bmatrix}
x_{c_i} & -X_{u_i} \\
\vdots & \vdots \\
\hline
y_{c_i} & -Y_{u_i} \\
\vdots & \vdots
\end{bmatrix}
\begin{bmatrix}
f \\
T_z
\end{bmatrix}
=
\begin{bmatrix}
K_{z_i} X_{u_i} \\
\vdots \\
\hline
K_{z_i} Y_{u_i} \\
\vdots
\end{bmatrix}.
$$

For the solution of this system to exist the calibration data cannot all lie in a single plane parallel to the sensor plane. Note that the solution to this system is an $f$ and $T_z$ that minimizes the error for the projection of the calibration data along the camera's $z$ axis. Ideally we would like to minimize the squared error for the projection of the calibration data in the camera's $xy$ image plane, as is done in most full camera calibration approaches. The advantage of the above approach is that it is direct rather than iterative. A less efficient but more consistent approach to estimating $f$, $C_x$, $C_y$ and $T_z$ is to use the values of the constant terms from the base lens setting in an iterative non-linear optimization for the four variable terms.

The final dynamic camera model for the given range of zoom consists of four variable terms $(f, C_x, C_y, T_z)$, and seven constant terms $(R_x, R_y, R_z, T_x, T_y, \kappa_1, s_x)$. Extending the model to a wider range of zoom positions would likely require modeling variations in

additional model terms (e.g. $\kappa_1$). To extend the camera model to additional control parameters (e.g. focus and aperture), the dimensionality of the calibration data would have to be increased. In addition, the dimensionality of any functions used to fit or to interpolate the values of the variable terms will also have to be increased.

## 3   Conclusions

The complex nature of the relationships between the control parameters and the imaging process plus the volume of calibration data and the range of conditions over which it must be taken combine to make the modeling and the calibration of cameras with automated zoom lenses fundamentally more difficult than that of cameras with fixed parameter lenses. To discover the degree of dependency between the terms of conventional models of the imaging process (such as the pinhole camera model) and the lens control parameters we need an iterative, empirical approach to modeling and calibration. The wide range of these dependencies can be exploited to reduce the amount of computation required to develop the model. With this approach we can efficiently produce a dynamic camera model that holds calibration across continuous ranges of control parameters.

## 4   Acknowledgments

## References

[1] M. Laikin. *Lens Design*. Marcel Dekker, New York, NY, 1991.

[2] K. Tarabanis, R. Y. Tsai, and D. S. Goodman. Modeling of a computer-controlled zoom lens. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 1545–1551, Nice, France, May 1992.

[3] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, RA-3(4):323–344, August 1987.

[4] A. G. Wiley. *Metric Aspects of Zoom Vision*. PhD thesis, University of Illinois at Urbana-Champaign, June 1991.

[5] R. G. Willson and S. A. Shafer. Precision imaging and control for machine vision research at Carnegie Mellon University. In *Proceedings of Conference on High-Resolution Sensors and Hybrid Systems*, volume 1656, pages 297–314, San Jose, CA, February 1992. SPIE.

[6] R. G. Willson and S. A. Shafer. What is the center of the image? In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, New York, NY, June 1993.